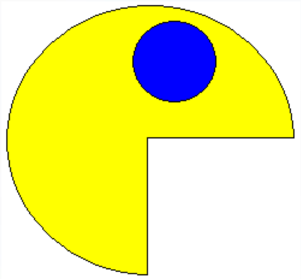
SMGT: Evolution by communication

Andrew Kuznetsov
Albert-Ludwig's University Freiburg, Germany
12.10.07



The idea

- Darwinian evolution is extremely slow. We have not so much time (about 3.6 billion years) to make an artificial design. But Lamarckian evolution and learning (Baldwin effect) could run faster
- Lateral gene transfer can accelerate the evolution too. Sperm mediated foreign gene transfer (SMGT) may be one of these mechanisms
- I suppose the model of artificial evolution and computation (Argo-machine) that should be faster than Darwinian evolution

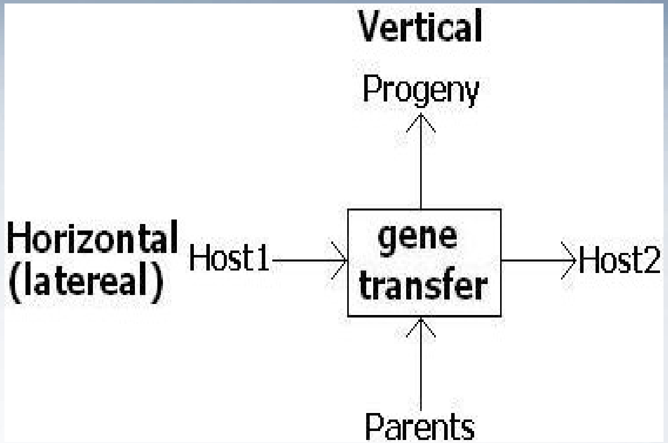
Sperm Mediated Gene Transfer: lost of stability, chaos and evolution?

Salvador Dali, 1944, **Dream Caused by…**

Kuznetsov, Kuznetsova, 1998, **Mobile Vector**

Vertical and horizontal gene transfer. Selfish DNA

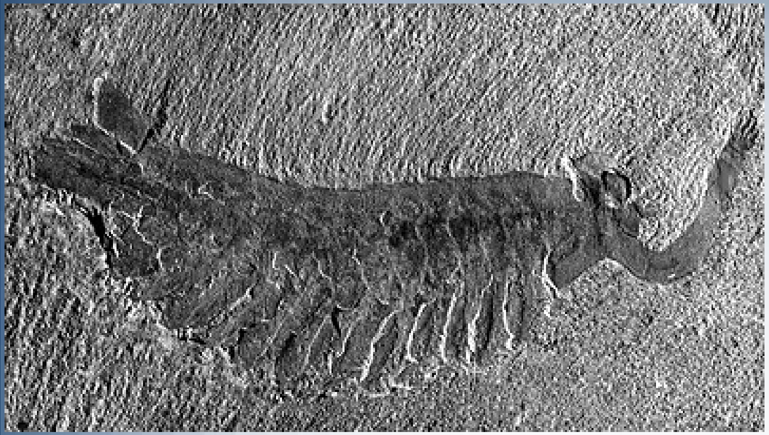
- Vertical inheritance within the host and horizontal transfer between hosts



Selfish DNA:

- Mobile DNA elements are widespread. They provide rearrangements in genes, which are exception to the general principle that genes are transmitted with great fidelity from parents to progeny. Rearrangements are typically rare, but are sometimes maintained by selective pressure. About 45% of human genome is composed of mobile DNA
- The canonical example of lateral transfer is movement of P-element from *Drosophila willistoni* to *D.melanogaster*, an event that might have occurred last century (Daniels et al., 1990)

Cambrian explosion, role of transpositions



- Cambrian explosion is an appearance of hard-bodied animals in fossil records ~530 million years ago. Over the next 70-80 million years, the rate of evolution accelerated by an order of magnitude
- What was the reason?
- *Opabinia regalis*, 4-8 cm, the Burgess Shale fauna, Cambrian – was an evolution experiment

Well established lateral gene transfer

Gene transfer from bacteria to bacteria:

- Transformation – gene transfer by uptake and incorporation of exogenous DNA (Griffith, 1928; Avery et al., 1944)
- Conjugation – a ‘sex’ process in bacteria (Lederberg, Tatum, 1946)
- Transduction – bacteriophage-mediated gene transfer (Morse et al., 1956)
- Competence – ability of cells to take up DNA, i.e. Gram-negative bacteria (*Haemophilus*, *Neisseria*, *Helicobacter* and *Acinetobacter*) as well as Gram-positive bacteria (*Bacillus*, *Mycobacterium* and *Streptomyces*)

- The uptake and stable maintenance of extracellular DNA – genetic transformation – is a major force in microbial evolution

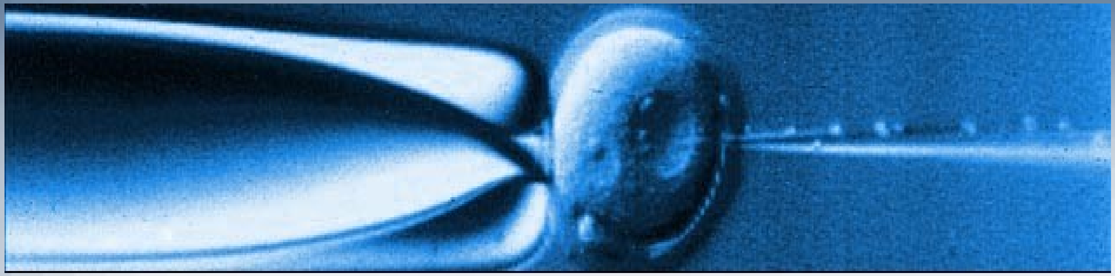
Transfer of octopine-type Ti-plasmid from *Agrobacterium tumefaciens* to plants results in tumor development to growth bacteria in infected plants (Montagu, Schell, 1982)

Lateral gene transfer into the human genome?

- The sequenced human genome contains 223 bacterial genes. Probably multiple independent gene transfers from different bacteria occurred (Lander et al., 2001)
- Some introduced genes appear to be involved in important physiological functions and have been fixed during evolution, because of the selective advantage they provide

- Howard M. Temin (1934-1994) proposed the retroviral origin of cancer (Rous Sarcoma Virus - RSV). In theory, retrovirus can transfer the oncogene from one individual to another that may lead to oncogenic transformation
- The available carrying capacity for retroviral vectors is ~7.5 kb (Verma, Somia, 1997), which is too small for most genes.
- Are there any others mechanisms for the lateral gene transfer into eukaryotes?

Bottle neck Effect, C-value Paradox, Red Queen Effect



- How can I transform a big multicellular organism? I have to introduce DNA into ovum or spermatozoon, on the single cell stage. How could a selfish DNA attack the genome? Possibly, the selfish DNA does so am I. It hits the sperm.

- C-value (Thomas, 1971) is a term for the DNA content of a cell. The greatest range of variations occur in unicellular eukaryotes: the range from yeast to amoeba is 80000-folds, and within green algae is 3000-fold. Why the amount of non-coding DNA is so great, and so variable?

- What about a relationship between the selfish DNA and the genome? Is it the arms raise or cooperation?

Brief history of sperm-mediated gene transfer

Brackett at al.	Uptake of heterologous genome by mammalian spermatozoa and its transfer to ova through fertilization	1971
Lavitrano et al.	Sperm cells as vectors for introducing foreign DNA into eggs: genetic transformation of mice	1989
Brinster at al.	No simple solution for making transgenic mice	1989
Khoo et al.	Sperm cells as vectors for introducing foreign DNA into zebrafish	1992
Tsai at al.	Sperm as a carrier to introduce an exogenous DNA fragment into the oocyte of Japanese abalone (*Haliotis divorsicolor suportexta)*	1997
Spadafora C.	Sperm cells and foreign DNA: a controversial relation	1998
Perry et al.	Mammalian transgenesis by intracytoplasmic sperm injection	1999
Lavitrano et al.	Efficient production by sperm-mediated gene transfer of human decay accelerating factor (hDAF) transgenic pigs for xenotransplantation	2002
Chang et al.	Effective generation of transgenic pigs and mice by linker based sperm-mediated gene transfer	2002

Sperm cells is an ideal vehicle for selfish DNA

- (e.g., HIV-1 binds with CD4 protein on the surface of sperm cells)
- Spermatozoa move to ova, which release an attractant. They compete for the ova. More active spermatozoa arrive at the ova
- SP from echinoids to man in certain conditions can take up foreign DNA. Motile SP capture the DNA better than nonmotile ones (Horan et al., 1991), but high DNA concentrations inhibit SP motility (Schit et al., 1998)
- Sperm nucleases are activated in response to the internalization of foreign DNA by sperm cells and cleave the DNA; the activity increases with the DNA concentration, i.e. critical amounts of DNA are 10 ng/10^6 epididymal mouse spermatozoa, 40 ng/10^6 epididymal boar spermatozoa, and 500 ng/10^6 ejaculated boar spermatozoa (Maione et al., 1997)

Expression of CMV-*lacZ* in loach *Misgurnus fossilis* fry after SMGT

Andreeva et al, 2003

- Control fry – mock analysis
- Experimental β-gal-positive fry 72 h after the eggs fertilization by sperm cells transfected with pcDNA3-*lacZ*

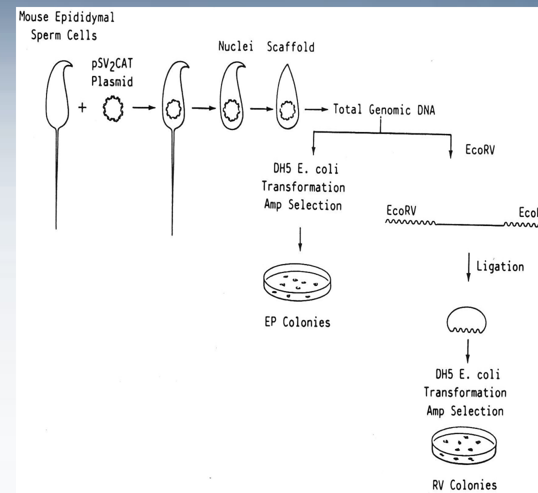
Only blue spots, dots and dashes are probably a result of transgene elimination during development

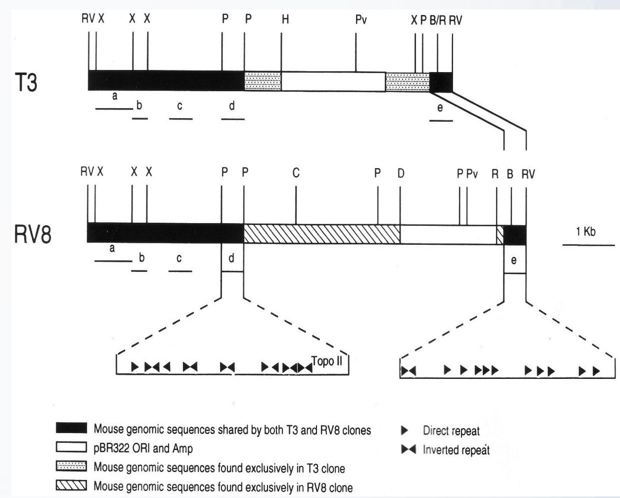
Sperm cells were squeezed out of the loach testis and intensively washed. Electric discharge (V=150 V, R=150 Ω, C=20 μF) was passed through the cell's suspension containing of 0.5 μg/ml DNA. Transfected sperm was added to eggs for fertilization. After development the embryos were fixed by 2.5% glutaraldehyde and stained by X-gal

Integration of foreign DNA into mouse sperm genome

Zoraqi, Spadafora, 1977

- pSV2*cat* is rescued from mouse sperm genome (EP – episome, RV – EcoRV digestion and ligation)
- Two mouse DNA sequences, identical in T3 and RV8 clones (solid), flank the unidentified DNA (dotted, hatched) within which the plasmid (open) has been integrated (Amp^R and *ori* of pSV2*cat*: 707-2542 in T3, 751-2620 in RV8). The TopoII consensus sequence adjacent to the one terminus of the site of integration; b,c,e – Alu-like repeats
- As a result of hybridization the same (solid) sequences were found in 14 randomly selected clones!





Conclusion for experiments

- Sperm cells take up DNA, giving them the double function of acting as a vehicle for transmitting not only their own but also foreign DNA
- Rescued plasmids were heavily rearranged, because sperm cells have enzymes, able to mediate DNA rearrangements
- The plasmids integrated into ’acceptor’ genomic site in the sperm DNA
- Random chromosomal DNA sequences appeared to integrate together with the plasmid DNA in the same genomic site
- Foreign DNA can be eliminated during development or not

Set of rules for SMGT

- Spermatozoon looks for ova by a chemical gradient
- The native DNA is transmitted by Mendelian rules
- Sperm cells can take up a foreign DNA (fDNA) from environment
- 2 rules for SP movement:
  - high amount of fDNA inhibits SP movement
  - low amount of fDNA activates its movement
- 2 rules for DNA integration:
  - high amount of fDNA leads to its fragmentation by sperm nucleases
  - low amount of fDNA doesn't activate those nucleases
- After a host died, DNA are fragmented and released into environment

- Discovered structures:
  - The DNA-uptake switch consists at least of 2 proteins, which bind to DNA, and includes 1 protein, which prevents DNA interaction
  - DNAs integrate into preferential sites of genome

SMGT extension

Corrado Spadafora, personal communication, 21 Sep 2005:

- "Our studies on SMGT greatly expanded in the last few years, looking essentially at the molecular basis of the process. We know now that SMGT is a retrotransposon-mediated process and that the sequences transferred from sperm cells to embryos, to adult animals and to the next generation are reverse transcribed copies of the original sequences incubated with spermatozoa. Reverse Transcriptase (RT) has a central role in the whole process.
- As a side product of these studies, we also found that RT is essential for preimplantaion embryonic development and has a role in human tumorigenesis: inhibition of RT arrest embryonic development and reduces the proliferation of human tumors."

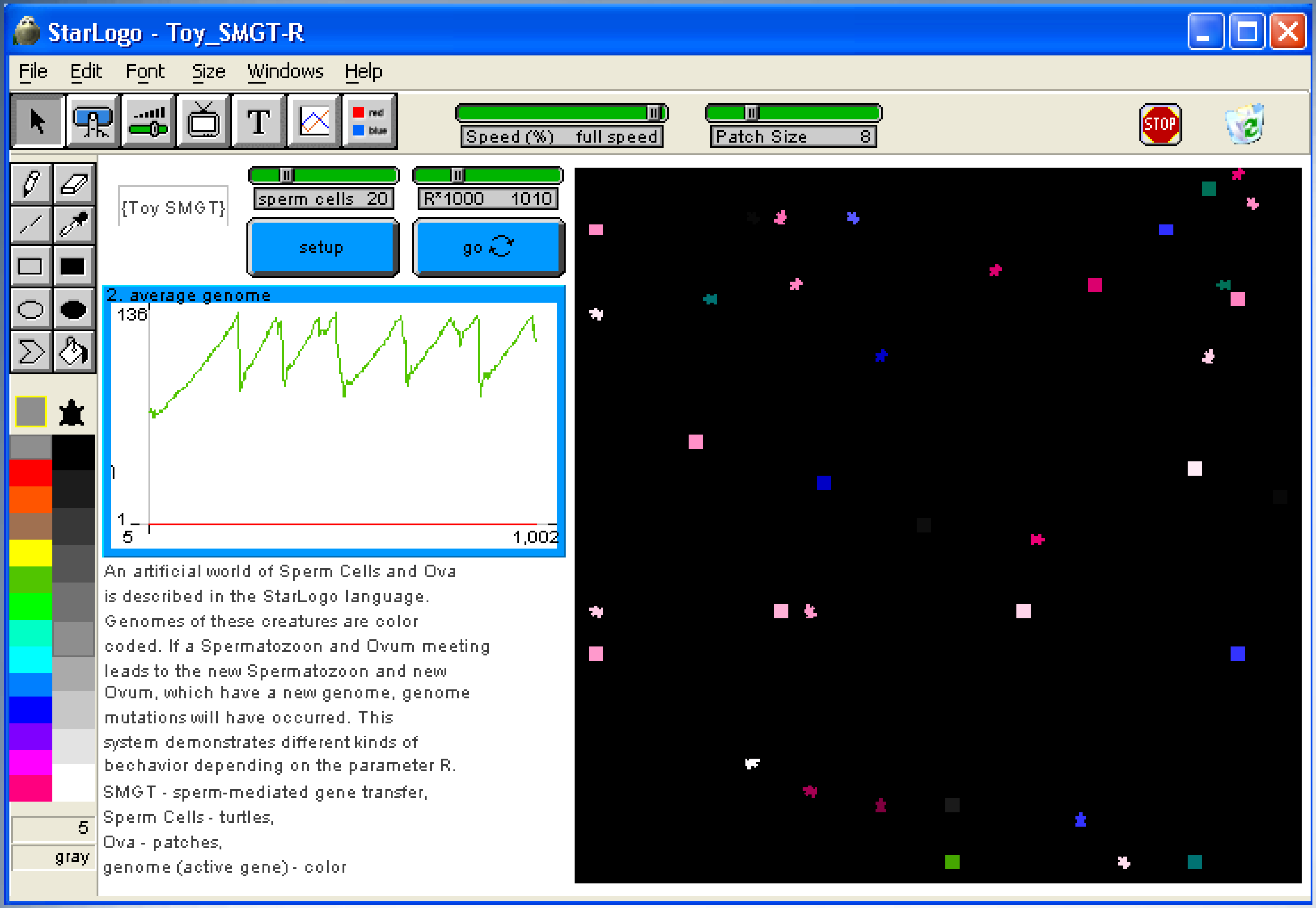
Applications

- *in theory*
  - formalization
- *in silico*
  - agent based simulation
- *in vitro*
  - bio-molecular computation
- *in vivo*
  - accelerated evolution

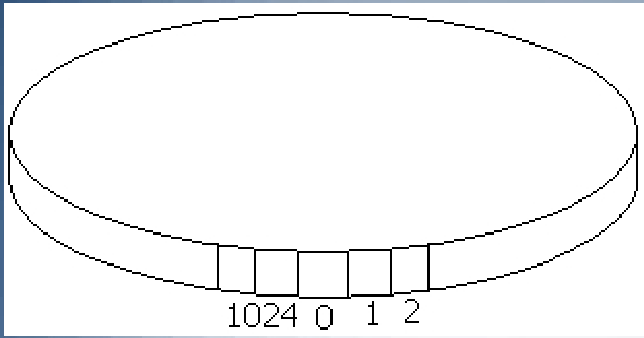
- Toward to the Artificial Life
  - easy conditions, minimum components, minimal set of rules

Agent based simulation

Kuznetsov, 2004



Toy model



Each creature has a circular genome consisting of 1024 'genes', only one of them is active and coded by color with mod(1024)

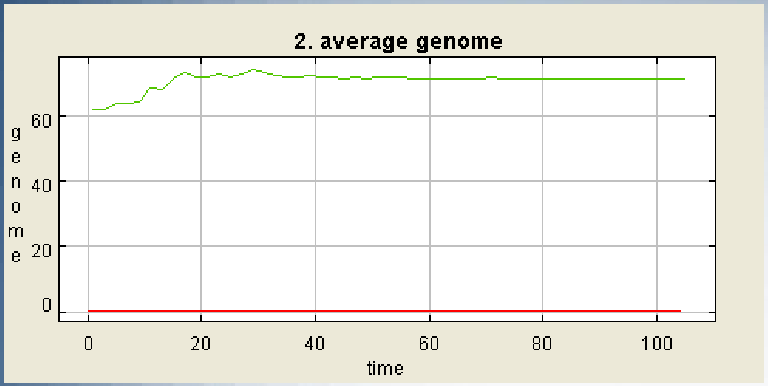
$$T(i+1) = |[T(i) + P(i)] / 2 * R|_{mod(1024)}$$

$$P(i+1) = T(i+1),$$

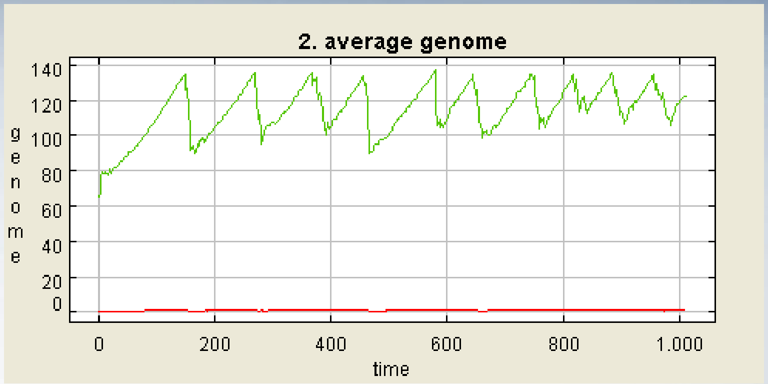
where **T(i)** is the color code of the individual Spermatozoon and **P(i)** is the color code of the individual Ovum at the time **i** of breeding. **R** is the mutation parameter on the interval]0, 4]

Behavior of the system

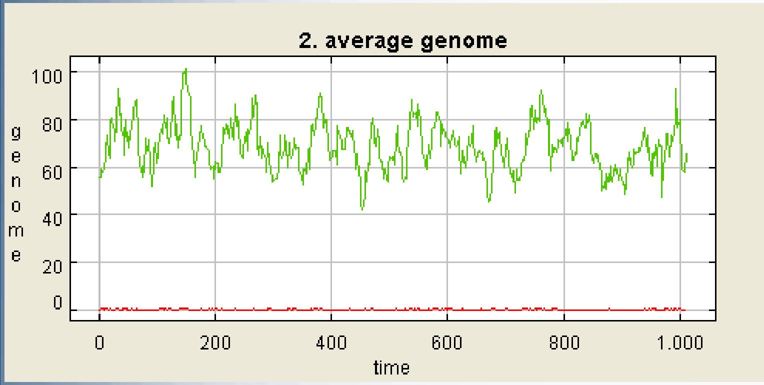
The system demonstrated ordered (R<=1) and complex (R>1) regimes



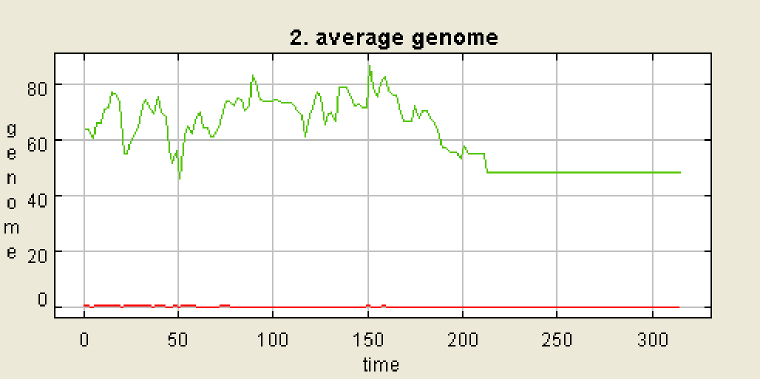
stable focus, R=1



periodic, R=1.01



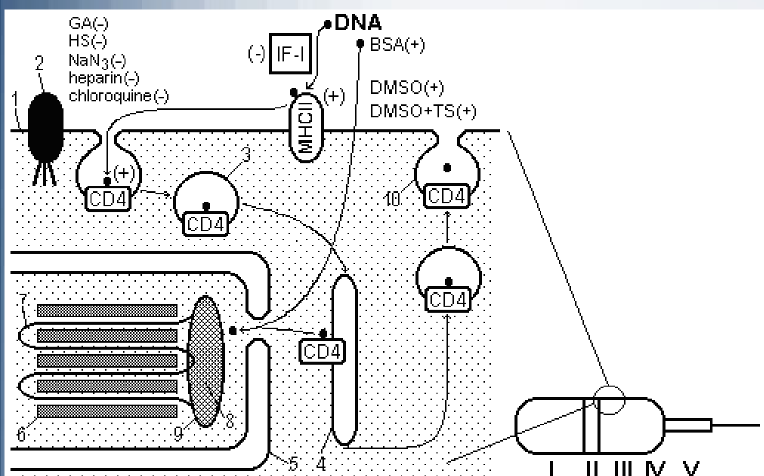
chaotic, R=3



strange attractor, R=4

SMGT is a possible accelerator of the biological evolution

Kuznetsov, Kuznetsova, 1995



Kuznetsov et al, 1998

DNA binding with the sperm surface

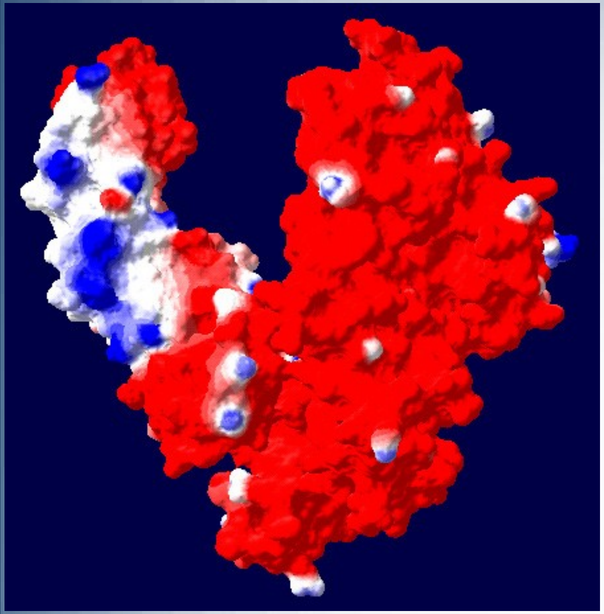
- control - block by heparin
- experiment - immunostaining

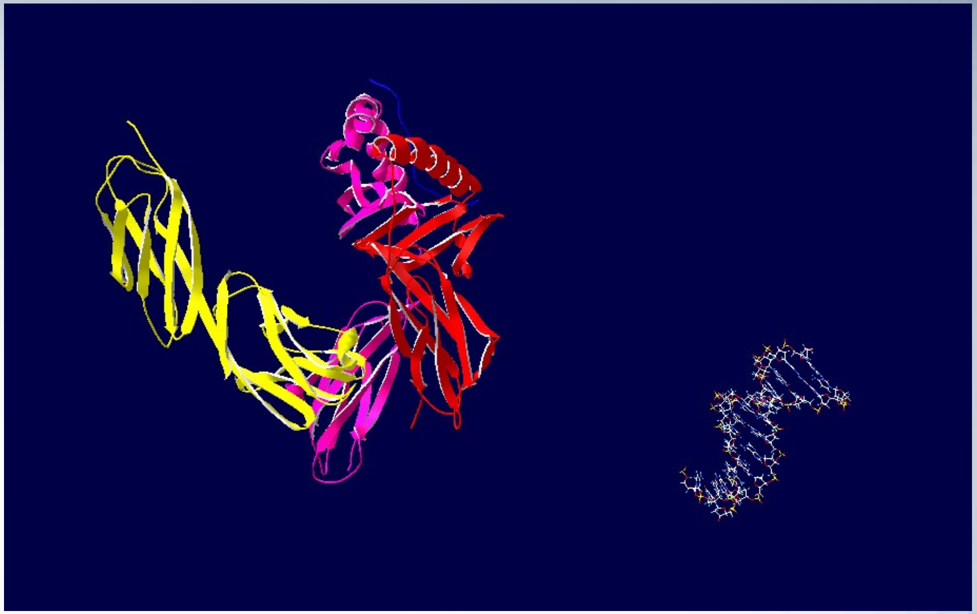
DNA penetration into spermatozoon

- 1 - plasmalemma, 2 - cytoskeleton bound protein, 3 - vesicle, 4 – releasing of DNA from vesicle into cytoplasm, 5 - nuclear envelope, 6 - protamines, 7 - chromosomal DNA loop, 8 - scaffold, 9 - nuclear annulus, 10 - releasing of DNA from vesicle out of cell; +/– - increase / decrease DNA uptake
- I - acrosomal cap, II - equatorial segment, III - postacrosomal region, IV - posterior ring, V - middle piece

CD4 and MHCII molecules are main actors on the scene

Lavitrano et al, 1997



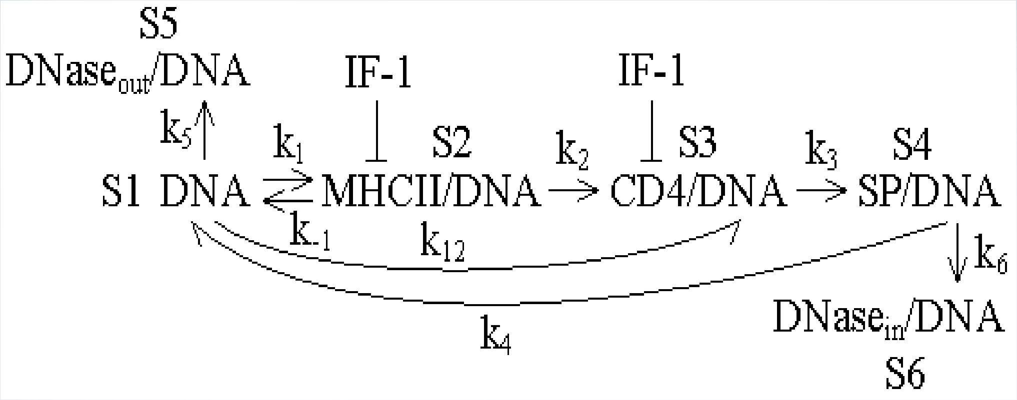


- But how does it work?
- Although the MHCII molecule has mostly a negative charged surface, it binds very strongly to DNA ($K_d = 7\times10^{-15}$ M, Wu et al, 1990)
- CD4 receptor introduces DNA into the sperm cell

Attacking the problem by ODE

Kuznetsov, 2005

Flow diagram of the foreign DNA processing in a sperm cell and the model



$dS1/dt = k_{-1}*S2 - (k_1 + k_5 + k_{12})*S1 + k_4*S4$

$dS2/dt = k_1*S1 - (k_{-1} + k_2)*S2$

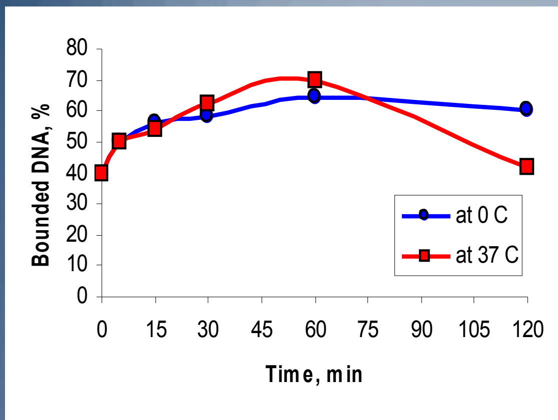
$dS3/dt = k_{12}*S1 + k_2*S2 - k_3*S3$

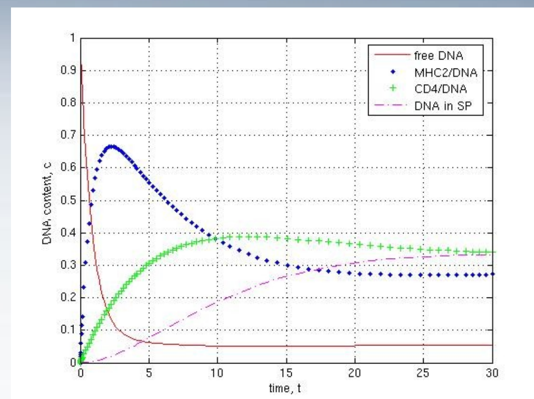
$dS4/dt = k_3*S3 - (k_4 + k_6)*S4$

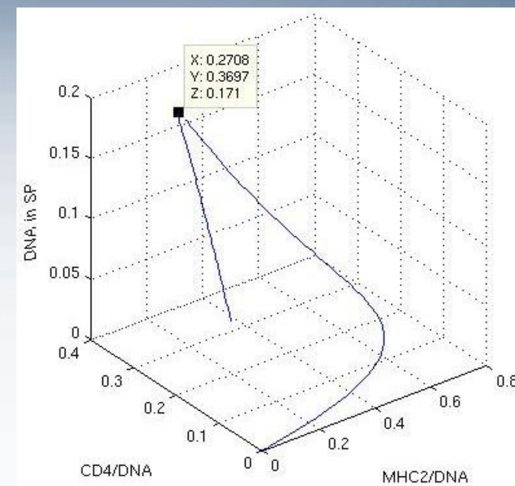
$dS5/dt = k_5*S1$

$dS6/dt = k_6*S4$

Experiment and model

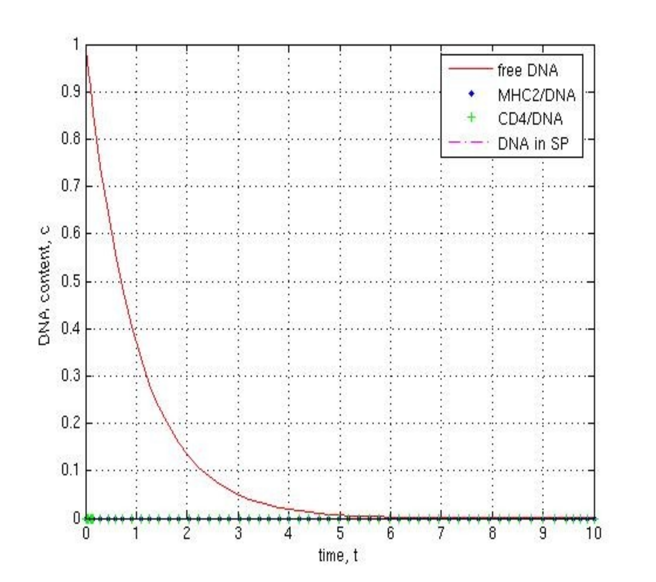


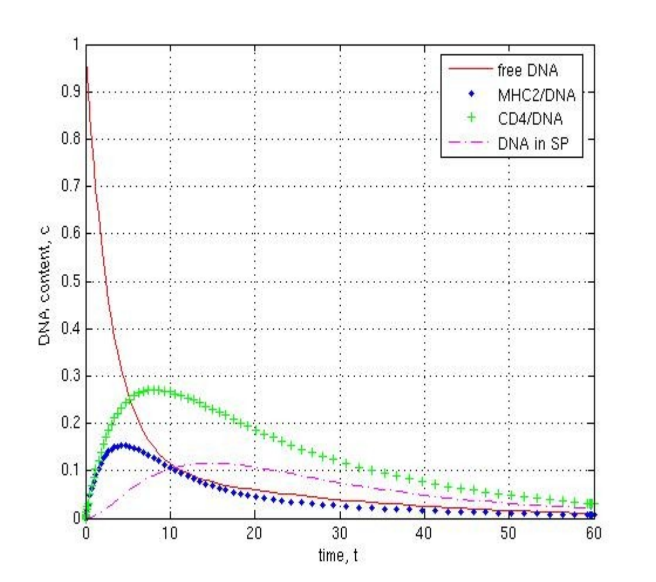




- **Interaction between radioactive DNA and real rabbit sperm cells** (Kuznetsov et al, 2000)
- **DNA interaction with an ideal 'electronic' sperm cell:** $k_1 = 1$, $k_{-1} = k_2 = k_{12} = k_3 = k_4 = 0.1$, $k_5 = k_6 = 0$. DNA binds very fast with MHCII proteins on the initial phase of reaction, interacts more slowly with CD4 receptors, penetrates into the sperm cell with a delay, and reaches the plateau at 30 t
- **3D phase diagram of bounded DNA:** *x* is MHCII/DNA, *y* is CD4/DNA, and *z* is SP/DNA. Three stages in the sperm/DNA interaction: 1) DNA binding with MHCII, 2) DNA interaction with CD4 and its internalization into sperm cell, 3) the slow digestion of DNA by an internal DNase ($k_6 = 0.1$), at $k_1 = 1$, $k_{-1} = k_{12} = k_2 = k_3 = k_4 = 0.1$, $k_5 = 0$

Different SMGT scenarios

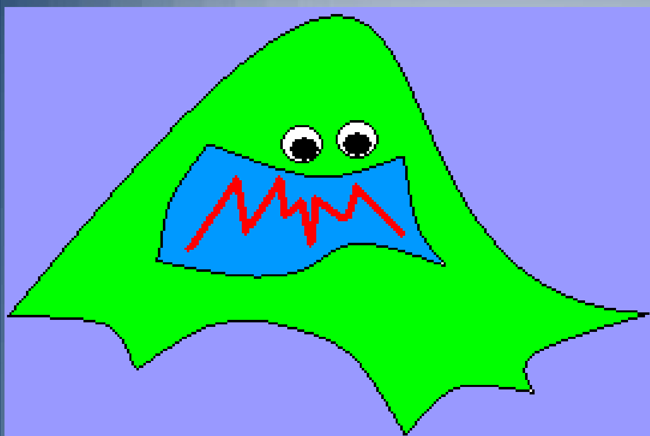


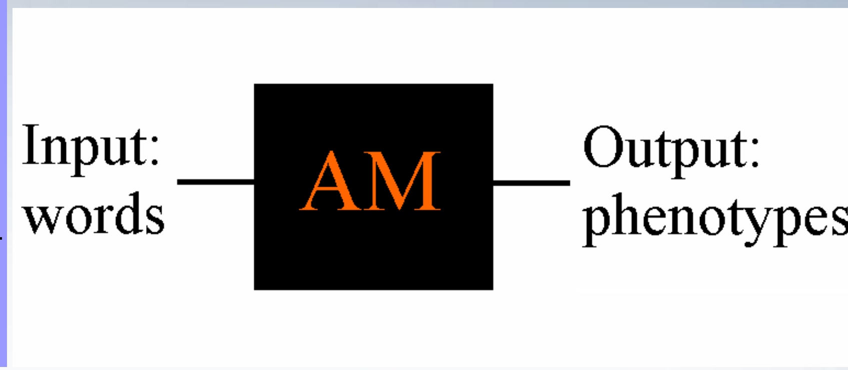


- **Protected conditions** (IF-1 and DNase activity in the seminal fluid: $k_1 = k_{12} = 0$, $k_5 = 1$), other coefficients were chosen $k_{-1} = k_2 = k_3 = k_4 = 0.1$, $k_6 = 0$
- **An example of the 'open' combination of parameters** ($k_1 = \ldots = k_6 = 0.1$), which leads to DNA penetration into sperm cell and possible genetic transformation

Argo-machine (a formal approach)

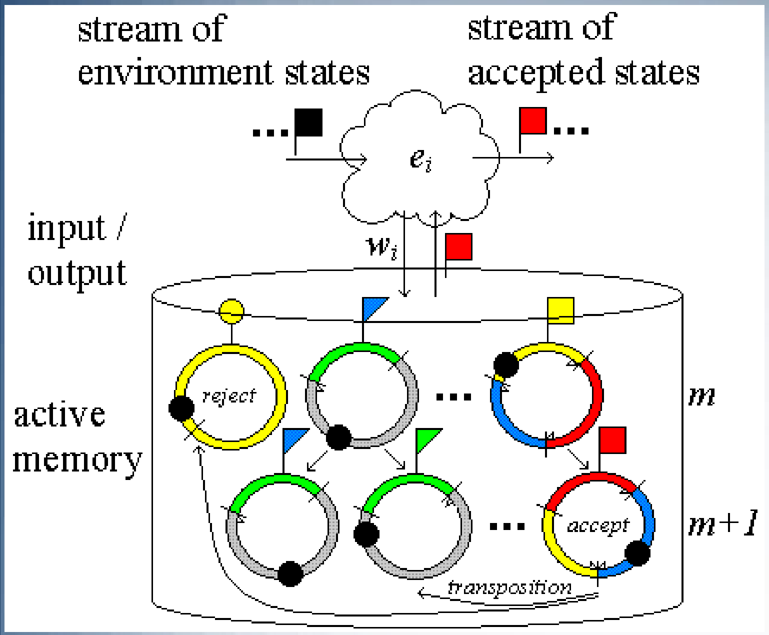
Kuznetsov et al, 2006





- “Consider an evolving system–an abstract machine and an environment that is continuously changing creates input words for the machine to stimulate an adaptation of this device to the surrounding…”
- ***AM*** is a non-deterministic abstract machine that searches according to oracle words in the design space to fit with its environment, cuts, transposes and pastes a set of tapes

Description



The system operates on inputs and memory, uploads the memory and yields outputs

- The *Argo-machine* (***AM***) consists of *agents*; each of these has a head, a tape and can be in different output states. The tape is a nonempty string of symbols that may be linear or circular. The head scans the tape according to an input word w_i, and cuts it at recognized sites. The agent arbitrarily pastes the tape. For each tape-configuration there is an appropriate output state of the agent that is checked by the environment. Special 'accept' and 'reject' states take immediate effect. An agent accepts, if its output state corresponds to the environment state; an agent will reject if less than two matches to the input word exist on the tape. ***AM*** can accept if at least one agent accepts, reject if all agents reject, or loop. If environment has changed, then it delivers a transposition and a new word w_{i+1}.
- The transposition means to make a copy of tape from the accepted agent to other ones and join it in head-to-tail
- ***AM*** looks for an agreement with the environment again and again

Argonaut algorithm

A* = "On word *w*:

1. Scan the tape to be sure that it contains at least two matches. If not, reject.
2. Cut at the matching sites and arbitrarily paste the tape's fragments.
3. Take the output state according the new tape.
4. Check it with the state of environment. If satisfy, accept; otherwise loop."

The computation

- Computation associated with *Argo-machine* is the shuffling of tapes from the initial set T_0 until an accept state is reached – an *adaptation*
- In general, the computation never ends, because the environment changes permanently; if it happens, the case, called as a *catastrophe*, leads to a transposition, generates a super-transition from the accept-state to the set of new initial-states, and brings a new generative word
- A progression of adaptations and catastrophes – an *evolution*

How does it work?

AM computation in winning branch

Language notations:

```
~,<,( - strings, cut before open brackets;
# - boundary symbol
```

Example 1. Adaptation without transposition:
environment '<~~>', word '<'

```
 1. <~~>        environment
 2. <           word
 3. #~<~<~<~#   tape_tick_1
 4. #~<~~><~#   tape_tick_2
 5. <~~>        accept
```

Example 2. Two adaptations with one transposition:
environment_1 '<~(~>', word_1 '<',
environment_2 '<~~~>', word_2 '('

```
 1. <~(~>                   environment_1
 2. <                       word_1
 3. #~(<~<~)<~#             tape_tick_1.1
 4. #~(<~(~><~#             tape_tick_1.2
 5. <~(~>                   accept_1
 6. <~~~>                   environment_2
 7. #~(<~<~)<~##~(<~(~><~#  transposition
 8. (                       word_2
 9. #~(<~<~)<~~(<~(~><~#    tape_tick_2.1
10. #~(<~<~)<~~~>)(~><~#    tape_tick_2.2
11. <~~~>                   accept_2
```

The elongation of input words leads to the increasing of building blocks

Alphabet: {a,b,c}
Language: {**a**,**ab**,**abc**}
Tape: aababcaabacbaa

Examples:

Case 1. On input word |**a**:
a **a**b **a**bc **a** **a**b **a**cb **a** **a**

Case 2. On input word |**ab**:
a **ab** **ab**ca **ab**acbaa

Case 3. On input word |**abc**:
aab **abc**aabacbaa

Description:

Case 1. Input is a short word; enormous number of rearrangements allows an exhaustive search, but all previous results are destroyed

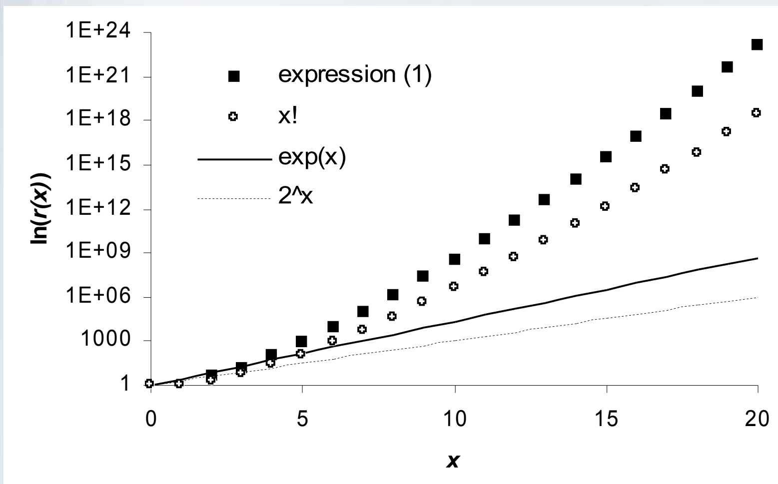
Case 2. **What language is optimal to maintain an appropriate level of diversity for a creative combinatorial design? What about the rules to form this language?**

Case 3. Input is a long word; deterministic kind of design

An analysis

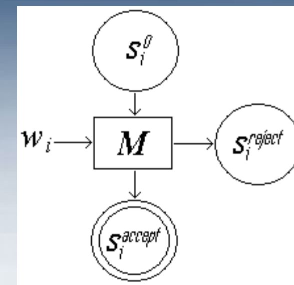
Combinatorial formula (1)

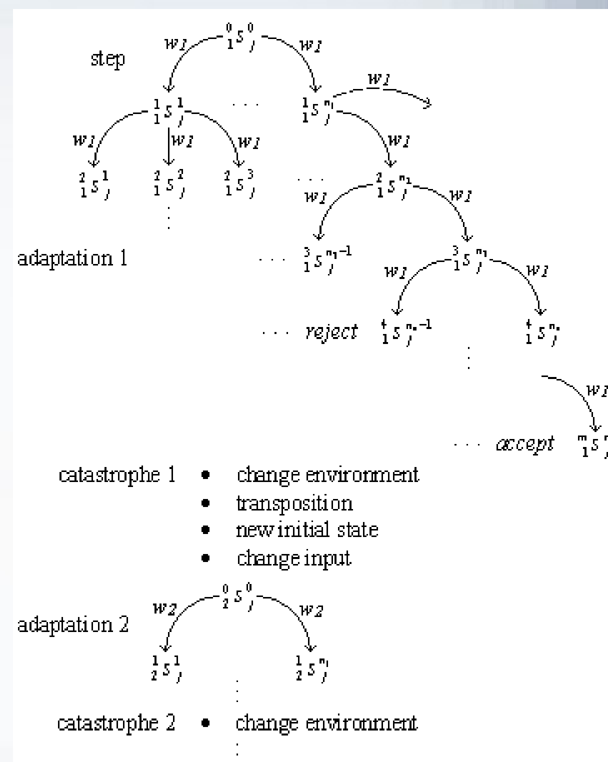
$$\begin{cases} r_0 = r_1 = 0, \\ r_x = 2^x * (x-1)!,\ x \geq 2 \end{cases} \quad (1)$$



Combinatorial power of expression (1)

Adaptation





Nondeterministic computation

Conclusion for the theory

- Complex behavior of Sperm and Ova emerged from the collective dynamics. The system demonstrated stable, periodic and chaotic regimes depending on the ‘mutation’ parameter

- The equilibrium state in ODE model, which corresponds to sperm/DNA block, is a special case. A large set of variations for parameters in the model caused 'leaky' states, where a part of exogenous DNA could penetrate sperm cells

- Argo-machine (***AM***) is a kind of constructive mutagenesis. ***AM*** is a set of agents, which act in parallel on their own tapes accordingly the instructions (input words), communicate each other by transpositions of the tapes and interact with an environment to compare own output states. The computation power of ***AM*** depends on the number of agents and the number of output states for each agent. The elongation of input words leads to bigger building blocks and to a hierarchical assembling

- SMGT could be seen as a global biological net for a spread of the selfish DNA. In addition, SMGT allows organisms to share genes. It is a source of innovations and the accelerator of evolution. SMGT can be even an inspiration for artificial communication systems

SMGT is a global net (references)

- *Kuznetsov A.V., Kuznetsova I.V.* Mobile vector. 1998. - 189 p. http://obomon.euro.ru/
- *Kuznetsov A.V., Pirkova A.V., Dvorianchikov G.A., Panfertsev E.A., Gavriushkin A.V., Kuznetsova I.V., Erokhin V.E.* Study of the transfer of foreign genes into mussel *Mytilus Galloprovincialis* Lam. eggs by spermatozoa. Russian Journal of Development Biology. 2001. V. 32(4). P. 309-318
- *Andreeva L.E., Sleptsova L.A., Grigorenko A.P., Gavriushkin A.V., Kuznetsov A.V.* Loach spermatozoa transfer foreign DNA, which expression is discovered in the early development stages. Russian Journal of Genetics. 2003. V. 39(6). P. 758-761
- *Kuznetsov A.V.* The rules of sperm-mediated gene transfer. Alife Mutants Hackingsession on Systems and Organisms (AMHSO), Rule 110 Winter Workshop, Bielefeld, Germany, 6-13 March 2004
- *Kuznetsov A.V.* Toy SMGT. ibid
- *Kuznetsov A.* DNA interaction with sperm cells: ODE model. BMC Systems Biology 2007, 1(Suppl 1):P42
- *Kuznetsov A., Schmitz M., Mueller K.* On Bio-Design of Argo-Machine. GWAL-7: 7th German Workshop on Artificial Life, Jena, Germany, 26-28 July 2006. P. 125-133